Developing Improved Algorithms for Detection and Analysis of Skin Cancer

By

Ammara Masood

Submitted in partial fulfilment of the requirements for the

Doctor of Philosophy

Faculty of Engineering and Information Technology

UNIVERSITY OF TECHNOLOGY, SYDNEY

July, 2016

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Acknowledgment

Over the last four years I have had the privilege of working with many people who have made my time at University of Technology, Sydney an enjoyable and intellectually stimulating experience.

I would like to thank all the people who have helped me along the way and contributed to this thesis. I am especially grateful to my Supervisor Associate Professor Dr. Adel Al-Jumaily. He has been very motivating throughout these years. While working with Dr. Adel, I have learned how to pursue my research with an intellectual rigor and how to critically evaluate my work and improve my research skills.

I would like to take this opportunity to acknowledge Scope Global for sponsoring my PhD through the Endeavour Post Graduate Award.

I would like to acknowledge Engr. Omama Masood for doing the proof reading and Engr. Tariq Adnan and Engr. Bani Kaur for helping in formatting corrections.

I would like to thank my friends and colleges who have encouraged and supported me to accomplish my degree.

Finally, I would like to acknowledge the endless love of my family who have been a constant source of support. Their faith in me and encouragement has helped me throughout my life.

Dedication

This thesis is dedicated to the five most important persons of my life without their support this journey would have been just a dream. Special thanks to my Grandfather M.R.A. Khan, my father M. Masood Alam Khan, my mother Azra Muzammil, my aunt Bushra Muzammil and my very best friend Tariq Adnan for being my inspiration and encouraging me throughout my career.

.

Abstract

Malignant melanoma is one of the deadliest forms of skin cancer and number of cases showed rapid increase in Europe, America, and Australia over the last few decades. Australia has one of the highest rates of skin cancer in the world, at nearly four times the rates in Canada, the US and the UK. Cancer treatment costs constitute more 7.2% of health system costs. However, a recovery rate of around 95% can be achieved if melanoma is detected at an early stage. Early diagnosis is obviously dependent upon accurate assessment by a medical practitioner. The variations of diagnosis are sufficiency large and there is a lack of detail of the test methods. This thesis investigates the methods for automated analysis of skin images to develop improved algorithms and to extend the functionality of the existing methods used in various stages of the automated diagnostic system. This in the long run can provide an alternative basis for researchers to experiment new and existing methodologies for skin cancer detection and diagnosis to help the medical practitioners.

The objective is to have a detailed investigation for the requirements of automated skin cancer diagnostic systems, improve and develop relevant segmentation, feature

selection and classification methods to deal with complex structures present in both dermoscopic/digital images and histopathological images.

During the course of this thesis, several algorithms were developed. These algorithms were used in skin cancer diagnosis studies and some of them can also be applied in wider machine learning areas. The most important contributions of this thesis can be summarized as below:

-Developing new segmentation algorithms designed specifically for skin cancer images including digital images of lesions and histopathalogical images with attention to their respective properties. The proposed algorithm uses a two-stage approach. Initially coarse segmentation of lesion area is done based on histogram analysis based orientation sensitive fuzzy C Mean clustering algorithm. The result of stage 1 is used for the initialization of a level set based algorithm developed for detecting finer differentiating details. The proposed algorithms achieved true detection rate of around 93% for external skin lesion images and around 88% for histopathological images.

-Developing adaptive differential evolution based feature selection and parameter optimization algorithm. The proposed method is aimed to come up with an efficient approach to provide good accuracy for the skin cancer detection, while taking care of number of features and parameter tuning of feature selection and classification algorithm, as they all play important role in the overall analysis phase. The proposed method was also tested on 10 standard datasets for different kind of cancers and results shows improved performance for all the datasets compared to various state-of the art methods.

- Proposing a parallelized knowledge based learning model which can make better use of the differentiating features along with increasing the generalization capability of the classification phase using advised support vector machine. Two classification algorithms were also developed for skin cancer data analysis, which can make use of both labelled and unlabelled data for training. First one is based on semi advised support vector machine. While the second one based on Deep Learning approach. The method of integrating the results of these two methods is also proposed. The experimental analysis showed very promising results for the appropriate diagnosis of melanoma. The classification accuracy achieved with the help of proposed algorithms was around 95% for external skin lesion classification and around 92 % for histopathalogical image analysis.

Skin cancer dataset used in this thesis is obtained mainly from Sydney Melanoma Diagnostic Centre, Royal Prince Alfred Hospital. While for comparative analysis and benchmarking of the few algorithms some standard online cancer datasets were also used. Obtained result shows a good performance in segmentation and classification and can form the basis of more advanced computer aided diagnostic systems. While in future, the developed algorithms can also be extended for other kind of image analysis applications.

Table of Contents

List of Abbreviations

Adaptive Differential evolution (ADE)

Adaptive Sequential Forward Floating Selection (ASFFS)

Adaptive Snake (AS)

Adaptive Thresholding based LS (AT-LS)

Advised Weighted Support Vector Machine (A-W-SVM)

Anisotropic Diffusion Filters (ADF)

Ant Colony Optimization (ACO)

Artificial Neural Network (ANN)

Atypical Pigment Network (APN)

Basal Cell Carcinoma (BCC)

Binary Particle Swarm Optimization (BPSO)

Circularity Index (CI)

Computer Aided Diagnosis (CAD)

Confocal Scanning Laser Microscopy (CSLM)

Derivatives of Gaussian (DOG)

Dice Similarity Coefficient (DSC)

Differential Evolution (DE)

Digital Dermoscopy Analysis (DDA)

Discrete Wavelet Transform (DWT)

Epiluminescence Microscopy (ELM)

Expectation Maximization (EM)

Expectation Maximization based level set (EM-LS)

Expertise Weight (EW)

False Negative Error (FNE)

False Positive Error (FPE)

Fast Fourier Transform (FFT)

Genetic Algorithm (GA)

Geodesic active Contour (GAC)

Gradient Vector Flow (GVF)

Gray Level Run length Matrix (GLRLM)

Gray Tone difference Matrix (GTDM)

Grey Level Co-Occurrence Matrix (GLCM)

Ground Truth (GT)

Hammoude Distance (HM)

Histogram analysis based Fuzzy C Mean thresholding (H-FCM)

Hue Saturation Lightness (HSL)

Hue-Saturation- Intensity (HSI)

Hybrid Genetic Algorithm (HGA)

Improved Binary Particle Swarm Optimization (IBPSO)

Independent Histogram Pursuit (IHP)

K Nearest Neighbour (KNN)

K mean clustering based LS (K-LS)

Karhunen-Loéve (KL)

Laplacian-of-Gaussian (LoG)

Level Set (LS)

Levenberg–Marquardt (LM)

Logistic regression (LogR)

Magnetic Resonance Imaging (MRI)

Malignant Melanoma (MM)

Multiple Layer Perceptron (MLP)

Optical Coherence Tomography (OCT)

Particle Swarm Optimization (PSO)

Peer Group Filter (PGF)

Pigmented Skin Lesions (PSLs)

Pigmented Spitz Naevi (PSNs)

Principal Component Analysis (PCA)

Radial Basis Function (RBF)

Receiver Operating Characteristic (ROC)

Restricted Boltzmann Machines (RBM)

Scaled Conjugate Gradient (SCG)

Segmented Region (SR)

Sequential Backward Elimination (SBE)

Sequential Floating Forward Selection (SFFS)

Sequential Forward Selection (SFS)

Spatial Fuzzy Clustering based LS (SF-LS)

Squamous Cell Carcinoma (SCC)

Statistical Region Merging (SRM)

Support Vector Machine (SVM)

Support Vectors (SV)

Total Body Photograph (TBP)

Total Dermatoscopy Score (TDS)

Transductive Support Vector Machines (TSVM)

True Positive Rate (TPR)

Typical Pigment Network (TPN)

Wavelet Packet Transform (WPT)

Wavelet Packet Transform (WPT)

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6